



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

52

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/091,472	03/07/2002	Jae Shin Yu	HI-0074	9044
34610	7590	04/06/2005	EXAMINER	
FLESHNER & KIM, LLP P.O. BOX 221200 CHANTILLY, VA 20153			TRAN, QUOC A	
			ART UNIT	PAPER NUMBER
			2176	

DATE MAILED: 04/06/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

## Office Action Summary

Application No.

10/091,472

Applicant(s)

YU ET AL.

Examiner

Quoc A. Tran

Art Unit

2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 07 March 2002.  
2a) ☐ This action is FINAL. 2b) ☒ This action is non-final.  
3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-20 is/are pending in the application.  
4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.  
5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.  
6) ☒ Claim(s) 1-20 is/are rejected.  
7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.  
8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.  
10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).  
11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).  
a) ☐ All b) ☐ Some \* c) ☒ None of:  
1. ☐ Certified copies of the priority documents have been received.  
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).  
\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)  
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)  
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date \_\_\_\_\_.  
4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date \_\_\_\_\_.  
5) ☐ Notice of Informal Patent Application (PTO-152)  
6) ☐ Other: \_\_\_\_\_.

### **DETAILED ACTION**

1. This action is responsive to application filed 03/07/2002, which was benefited from foreign priority filed 03/09/2001.
2. Claims 1-20 are pending. Claims 1, 13 and 17 are independent claims.

### ***Priority***

3. Acknowledgment is made of applicant's claim for foreign priority based on an application filed in Republic Of Korea on 03/09/2001. It is noted, however, that applicant has not filed a certified copy of the foreign priority application as required by 35 U.S.C. 119(C). Clarification and/or correction are required.

### ***Claim Rejections - 35 USC § 102***

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

**Claims 17-20 are rejected under 35 U.S.C. 102(e) as being anticipated by Gibbon et al. US006714909B1 filed 11/21/2000 (hereinafter '909).**

**In regard to independent claim 17,** *"determining the sizes of weight determining factors; determining weights based upon the sizes of the weight determining factors; and adding values obtained by multiplying the weight determining factors with corresponding weights", as taught by '909 at col. 8, lines 45-60 (i.e... A GMM model consists of a set of weighted Gaussian... where  $K$  is the number of mixtures,  $M_{sub.i}$  and  $\Sigma_{sub.i}$  are the mean vector and covariance matrix of the  $i$ th mixture, respectively, and  $\omega_{sub.i}$  is the weight associated with the  $i$ th Gaussian. Based on training data, the parameter set  $\lambda=(\omega, M, \Sigma)$  is optimized such that  $f(x)$ (best fits the given data....). Examiner reads weighted Gaussian formula, wherein  $M_{sub.i}$  and  $\Sigma_{sub.i}$  are the mean vector and covariance matrix of the  $i$ th mixture, respectively, and  $\omega_{sub.i}$  is the weight associated with the  $i$ th Gaussian, which could interpreted as claimed.*

**In regard to dependent claim 18,** *"wherein the weight determining factors include the size of the text areas, mean text size in the text area and the display duration time of a text", as taught by '909 at col. 13, lines 30-50 (i.e... FIG. 14 is a window that plays back streaming content to a user. It is triggered when users click on a particular item. In this playback window, the upper portion shows the video and the lower portion the text synchronized with the video. During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given...).*

**In regard to dependent claim 19**, *"wherein the mean text size in the text area is determined by the densities and sizes of histograms about the text area"*, as taught by '909 at col. 13, line 60 through col. 14, line 5 (i.e... Within the boundary of each story, a keyword histogram is first constructed as shown in FIG. 15 ....A fixed number of key frames within the boundary are chosen so that they (1) are not within anchor speech segments and (2) yield maximum covered area with respect to the keywords histogram. The peak points marked on the histogram in FIG. 15 indicate the positions of the chosen frames and the shaded area underneath them defines the total area coverage on the histogram by the chosen key frames...).

**In regard to dependent claim 20**, *"wherein the display duration time of the text is determined by considering whether a previously extracted text area is identical to a currently extracted text area"*, as taught by '909 at col. 10, lines 50-65 (i.e... block of text available at this point, the task is to determine how these blocks of text can be merged to form semantically coherent content based on appropriate criteria. Since news introductions are to provide a brief and succinct message about the story, they naturally have a much shorter duration than the detailed news reports. Based on this observation, in step 5060, a headline story segmentation unit 440 initially classifies each block of text as a story candidate or an introduction candidate based on duration. ...), also as taught by '909 at col. 12, lines 15-25 (i.e... blocks formed in this way not only contain enough information for similarity comparison but also have natural breaks of chains of repeated words if true boundaries are present...).

***Claim Rejections - 35 USC § 103***

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

**Claims 1-16 are rejected under 35 U.S.C. 103(a) as being unpatentable by Gibbon et al. US006714909B1 filed 11/21/2000 (hereinafter '909), in view of Nelson et al. US006243713B1 – filed 08/24/1998 (hereinafter '713).**

**In regard to independent claim 1, “*extracting a plurality of text areas from a video stream; calculating importance measures according to weights for each of the extracted text areas*”, as taught by '909 at col. 2, lines 1-30 (i.e... ability to segment multimedia data, such as news broadcasts, into retrievable units that are directly related to what users perceive as meaningful... separating a multimedia data stream into audio, visual and text components, segmenting the audio, visual and text components based on semantic differences...),**

**“*and synthesizing the text areas to be synthesized into the key frame*” as taught by '909 at col. 13, lines 30-40 (i.e... FIG. 14 is a window that plays back streaming content to a user. It is triggered when users click on a particular item. In this playback window, the upper portion shows the video and the lower portion the text synchronized**

Art Unit: 2176

with the video. During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time...);

'909 does not explicitly teach, *"selecting the number of text areas to be synthesized based upon the importance measures in the order of higher importance"*, however as taught by '713 at col. 6, lines 5-50 (i.e...Compound documents are separated 110 into constituent multimedia components of different data types, such as text, images, video, audio/voice, and other data types...portion thereof) that is associated with the token, and may include the actual, or preferably, processed data extracted from, and representative of the original component..."compound") query 150 specified by the user, which may have one or more multimedia components (such as text portions, image portions, video portions, or audio portions). Preferably these various multimedia components are combined with one or more query operators...includes both text 151 and image 157 components, and a number of query operators 152 defining both logical relationships 152 and proximity relationships 156 between the multimedia components...).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified '713 into '909 to provide a way, wherein selecting the number of text areas to be synthesized based upon the importance measures in the order of higher importance. One of the ordinary skills in the art would have been motivated to perform such a modification to provide a desirable system that retrieves compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video, as taught

by '713 at col. 2, lines 10-20 (i.e....retrieves compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video).

**In regard to dependent claim 2,** *"wherein the text areas are extracted according to certain intervals of the video stream"*, as taught by '909 at col. 14, lines 15-35 (i.e... During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given...).

**In regard to dependent claim 3,** *"wherein the synthetic key frame is generated in each of the certain intervals of the video stream"*, as taught by '909 at col. 13, lines 15-35 (i.e... During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given...).

**In regard to dependent claim 4,** *"wherein the certain intervals of the video stream are discriminated by scenes as logical edition units of a video"*, as taught by '909 at col. 14, lines 40-50 (i.e... FIG. 16 is the visual representation about a story on El Nino and FIG. 17 is the visual representation of a story about the high suicide rate among Indian youngsters in a village. As can be seen from both these figures that the story representations constructed this way are compact, semantically revealing, and visually informative with respect to the content of the corresponding stories. A user can choose either to scroll the text on the right to read the story or to click on the button of that story in the table of contents to playback synchronized audio, video, and text, all



starting from where the story begins. A different alternative maybe to click on one of the representative images to playback multimedia content starting from the point of time where the chosen image is located in the video. Compared with linear browsing or low level scene cut browsing, this system allows a more effective content based non-linear information retrieval...). Examiner reads linear browsing or low level scene cut browsing, which could interpreted as claimed, "...scenes as *logical edition units of a video...*", see specification at the Background Of The Invention, page 1, lines15-20 (i.e... The most basic technique for a non-linear video content browsing and searching is a shot segmentation scheme and a shot clustering scheme, both of which are the most critical for structurally analyzing multimedia contents...).

**In regard to dependent claim 5, wherein the certain intervals of the video stream are discriminated by shots as physical edition units of a video",** as taught by '909 at col. 14, lines 40-50 (i.e... FIG. 16 is the visual representation about a story on El Nino and FIG. 17 is the visual representation of a story about the high suicide rate among Indian youngsters in a village. As can be seen from both these figures that the story representations constructed this way are compact, semantically revealing, and visually informative with respect to the content of the corresponding stories. A user can choose either to scroll the text on the right to read the story or to click on the button of that story in the table of contents to playback synchronized audio, video, and text, all starting from where the story begins. A different alternative maybe to click on one of the representative images to playback multimedia content starting from the point of time where the chosen image is located in the video. Compared with linear browsing or low

level scene cut browsing, this system allows a more effective content based non-linear information retrieval...). Examiner reads linear browsing or low level scene cut browsing, which could interpreted as claimed, "...scenes as *logical edition units of a video...*", see specification at the Background Of The Invention, page 1, lines 15-20 (i.e... The most basic technique for a non-linear video content browsing and searching is a shot segmentation scheme and a shot clustering scheme, both of which are the most critical for structurally analyzing multimedia contents...).

**In regard to dependent claim 6, "wherein the weights are determined in proportion to the size of the text area, the mean text size of the text area and the display duration time of a text",** as taught by '909 at col. 13, lines 30-50 (i.e... FIG. 14 is a window that plays back streaming content to a user. It is triggered when users click on a particular item. In this playback window, the upper portion shows the video and the lower portion the text synchronized with the video. During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In the black box at the bottom, the timing with respect to the starting point of the program is given...).

**In regard to dependent claim 7, "wherein the mean text size in the text area is determined by using the density and size of a histogram for the text area",** as taught by '909 at col. 13, line 60 through col. 14, line 5 (i.e... Within the boundary of each story, a keyword histogram is first constructed as shown in FIG. 15 ...A fixed number of key frames within the boundary are chosen so that they (1) are not within anchor speech segments and (2) yield maximum covered area with respect to the keywords histogram.

The peak points marked on the histogram in FIG. 15 indicate the positions of the chosen frames and the shaded area underneath them defines the total area coverage on the histogram by the chosen key frames...).

**In regard to dependent claim 8**, *"wherein the display duration time of the text is determined by considering whether a previously extracted text area is identical to a currently extracted text area"*, as taught by '909 at col. 10, lines 50-65 (i.e... block of text available at this point, the task is to determine how these blocks of text can be merged to form semantically coherent content based on appropriate criteria. Since news introductions are to provide a brief and succinct message about the story, they naturally have a much shorter duration than the detailed news reports. Based on this observation, in step 5060, a headline story segmentation unit 440 initially classifies each block of text as a story candidate or an introduction candidate based on duration. ...), also as taught by '909 at col. 12, lines 15-25 (i.e... blocks formed in this way not only contain enough information for similarity comparison but also have natural breaks of chains of repeated words if true boundaries are present...).

**In regard to dependent claim 9**, *"wherein the weight increases as the size of the text area, the mean text size in the text area or the display duration time of the text increases"*, as taught by '909 at col. 13, lines 30-50 (i.e... FIG. 14 is a window that plays back streaming content to a user. It is triggered when users click on a particular item. In this playback window, the upper portion shows the video and the lower portion the text synchronized with the video. During playback, audio is synchronized with video. Either key frames or the original video stream is played back. The text scrolls up with time. In

the black box at the bottom, the timing with respect to the starting point of the program is given... keywords are chosen in step 5080 above, from the story according to their importance computed as weighted frequency).

**In regard to dependent claim 10**, *"wherein the number of the text areas to be synthesized is selected from the plurality of text areas in the order of importance"*, however as taught by '713 at col. 6, lines 5-50 (i.e...Compound documents are separated 110 into constituent multimedia components of different data types, such as text, images, video, audio/voice, and other data types...portion thereof) that is associated with the token, and may include the actual, or preferably, processed data extracted from, and representative of the original component..."compound") query 150 specified by the user, which may have one or more multimedia components (such as text portions, image portions, video portions, or audio portions). Preferably these various multimedia components are combined with one or more query operators...includes both text 151 and image 157 components, and a number of query operators 152 defining both logical relationships 152 and proximity relationships 156 between the multimedia components...).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified '713 into '909 to provide a way, wherein the number of the text areas to be synthesized is selected from the plurality of text areas in the order of importance. One of the ordinary skills in the art would have been motivated to perform such a modification to provide a desirable system that retrieves compound documents in response to queries that include various multimedia elements

in a structured form, including text, image features, audio, or video, as taught by '713 at col. 2, lines 10-20 (i.e...retrieves compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video).

**In regard to dependent claim 11,** *“wherein the number the text areas to be synthesized is determined according to browser size”*, as taught by '909 at col. 3, lines 5-10 (i.e... FIG. 14 illustrates the representation of a playback interface ... FIG. 15 illustrates a histogram of keywords within a story).

**In regard to dependent claim 12,** *“wherein the sizes of the text areas to be synthesized are determined according to browser size”*, as taught by '909 at col. 3, lines 5-10 (i.e... FIG. 14 illustrates the representation of a playback interface ... FIG. 15 illustrates a histogram of keywords within a story).

**In regard to independent claim 13,** incorporate substantially similar subject matter as cited in claim 1 above, and is similarly rejected under the same rationale.

**In regard to dependent claim 14,** incorporate substantially similar subject matter as cited in claim 6 above, and is similarly rejected under the same rationale.

**In regard to dependent claim 15,** *“wherein the certain rule is addition of values obtained by multiplying the weight determining factors with the corresponding weights”*, as taught by '909 at col. 8, lines 45-60 (i.e... A GMM model consists of a set of weighted Gaussian... where  $K$  is the number of mixtures,  $M_{sub.i}$  and  $\Sigma_{sub.i}$  are the mean vector and covariance matrix of the  $i$ th mixture, respectively, and  $\omega_{sub.i}$  is the weight associated with the  $i$ th Gaussian. Based on training data, the parameter set

Art Unit: 2176

$\lambda = (\omega, M, \Sigma)$  is optimized such that  $f(x)$  (best fits the given data....).

Examiner reads weighted Gaussian formula, wherein  $M_{sub.i}$  and  $\Sigma_{sub.i}$  are the mean vector and covariance matrix of the  $i$ th mixture, respectively, and  $\omega_{sub.i}$  is the weight associated with the  $i$ th Gaussian, which could interpreted as claimed.

**In regard to dependent claim 16, “wherein the number of the text areas to be synthesized is selected from the plurality of text areas in the order of importance”,**

however as taught by '713 at col. 6, lines 5-50 (i.e...Compound documents are separated 110 into constituent multimedia components of different data types, such as text, images, video, audio/voice, and other data types...portion thereof) that is associated with the token, and may include the actual, or preferably, processed data extracted from, and representative of the original component...“compound”) query 150 specified by the user, which may have one or more multimedia components (such as text portions, image portions, video portions, or audio portions). Preferably these various multimedia components are combined with one or more query operators...includes both text 151 and image 157 components, and a number of query operators 152 defining both logical relationships 152 and proximity relationships 156 between the multimedia components...).

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to have modified '713 into '909 to provide a way, wherein the number of the text areas to be synthesized is selected from the plurality of text areas in the order of importance. One of the ordinary skills in the art would have been motivated to perform such a modification to provide a desirable system that retrieves

Art Unit: 2176

compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video, as taught by '713 at col. 2, lines 10-20 (i.e...retrieves compound documents in response to queries that include various multimedia elements in a structured form, including text, image features, audio, or video).

### ***Conclusion***

6. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Gibbon	US006473778B1	filed 02/01/1999
--------	---------------	------------------

Dimitrova	US006363380B1	filed 01/13/1998
-----------	---------------	------------------

7. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Quoc A. Tran whose telephone number is (571) 272-4103. The examiner can normally be reached on Monday through Friday from 8:30AM to 5:00PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph H. Feild can be reached on (571) 272-4090. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for

Art Unit: 2176

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see <http://pair-direct.uspto.gov>.

Should you have questions on access to the Private PAIR system, contact the


Electronic Business Center (EBC) at 866-217-9197 (toll-free).

**Quoc A. Tran**

**Patent Examiner**

**Technology Center 2176**

**March 30, 2005**



**SANJIV SHAH**  
**PRIMARY EXAMINER**